Environment and Planning B: Planning and Design 2002, volume 29, pages 841 – 865

DOI:10.1068/b12813

Urban settlement transitions

Claes Andersson
EES-6 MS T003, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, and Physical Resource Theory, Göteborg University/Chalmers University of Technology, 412 96 Göteborg, Sweden; e-mail: andersson@lanl.gov, claesand@fy.chalmers.se

Steen Rasmussen
EES-6 MS T003 and T-CNLS MS B258, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA; e-mail: steen@lanl.gov, steen@santafe.edu

Roger White
Department of Geography, Memorial University, St. John's, Newfoundland A1B 3X9, Canada, and RIKS, Postbox 463, 6200 AL Maastricht, The Netherlands; e-mail: roger@plato.ucs.mun.ca, roger@riks.nl
Received 6 March 2001; in revised form 22 December 2001

Abstract. Urban growth dynamics attracts the efforts of scientists from many different disciplines with objectives ranging from theoretical understanding to the development of carefully tuned realistic models that can serve as planning and policy tools. Theoretical models are often abstract and of limited applied value while most applied models yield little theoretical understanding. Here we present a mathematically well-defined model based on a modified Markov random field with lattice-wide interactions that produces realistic growth patterns as well as behavior observed in a range of other models based on diffusion-limited aggregation, cellular automata, and similar models. We investigate the framework's ability to generate plausible patterns using minimal assumptions about the interaction parameters since the tuning and specific definition of these are outside of the scope of this paper. Typical universality classes of the simulated dynamics and the phase transitions between them are discussed in the context of real urban dynamics. Using suitability data derived from topography, we produce configurations quantitatively similar to real cities. Also, an intuitive class of interaction rules is found to produce fractal configurations, not unlike vascular systems, that resemble urban sprawl. The dynamics are driven by interactions, depicting human decisions, between all lattice points. This is realized in a computationally efficient way using a mean-field renormalization (area average) approach. The model provides a mathematically transparent framework to which any level of detail necessary for actual urban planning application can be added.

1 Introduction

Urban growth modeling has evolved over recent years to capture increasingly well the details of urban morphology and structure on a qualitative as well as a quantitative level. In this paper we are concerned mainly with demonstrating the ability of the proposed framework to produce realistic urban growth patterns. Urban growth is a complex dynamical process in which many of the patterns which are observed at a macroscopic scale emerge as a result of microscopic dynamics. It is also often macroscopic properties that can be directly observed and measured in real systems. Therefore, comparative analysis of patterns observed in simulation and reality can be helpful for validating a model and thereby gaining knowledge about microscopic properties. Fractal analysis has been used extensively in the analysis of urban growth modeling and can capture morphological properties that elude other means of measurement. In addition to fractal dimension analysis we also introduce complementary methods and concepts for spatial analysis that are borrowed from statistical physics.

Broadly speaking, urban modeling is directed at two objectives. On the one hand, modeling is aimed at providing a deeper theoretical understanding of urban dynamics.

On the other hand, effort is being directed at developing realistic, applied models that can serve as planning and policy tools. In the past these two goals have not been easily reconciled. Classical theoretical models, such as the Alonso – Muth land-use model (Alonso, 1964) and its descendants in the new urban economics (Papageorgiou, 1990), lack realism. Entropy-maximizing models and related approaches based on spatial interaction models developed in the 1960s and 1970s (Batty, 1976; Wilson, 1970) were much more realistic. They were shown to be capable of generating spatial structure, and are in a sense similar to the approach developed here—for example, both assume that a spatial structure and spatial interaction are mutually determined. However, the emphasis was primarily on predicting interzonal flows rather than achieving a theoretical understanding of spatial structure, although White (1977; 1978) used the approach to identify a bifurcation in spatial structure that depends on interaction parameters.

The advent of new modeling techniques has at least permitted the formulation of models that are both realistic and theoretically rich. Examples of these techniques, ordered roughly from the more theoretical to the more applied, are: sociodynamic (Weidlich, 1995; 2000), diffusion-limited aggregation (DLA) (Batty and Longley, 1994; Witten and Sander, 1983), correlated percolation (Makse et al, 1998), cellular automata (CA) models (Batty and Xie, 1994; Benguigui, 1995; Benguigui and Daoud, 1991; Couclelis, 1997; Engelen et al, 1997; Takeyama, 1996; White et al, 1997; Wu, 1998a; Xie, 1996), agent-based (Sanders et al, 1997) and integrated agent-based – CA models (Benenson and Portugali, 1997; Portugali, 2000), and integrated CA – GIS models (Candau et al, 2000; Clarke and Gaydos, 1998; Clark et al, 1997; Wu, 1998b). Although particular models developed using these techniques may be designed to emphasize theory over realism or vice versa, in every case the models not only have large-scale theoretical implications but also produce rich detail.

The design of the framework that we propose in this paper grew from the realization that certain properties are likely to be common to many different classes of systems that incorporate spatial interactions. Pattern formation is often surprisingly robust to changes in the details of the rules governing its dynamics and many important aspects of physical models are not depicting the phenomena being modeled specifically but more generally model how interactions take place between points in a space. By no means are we the first to realize these relations; they are in fact central themes behind well-established concepts originating from natural science, such as universality. They have also been applied to social sciences and urban systems (see, for example, Portugali, 2000; Weidlich, 2000). Also, maximum-entropy models (to which our model is related through its use of the Maxwell – Boltzmann distribution) dates back to the 1950s in the context of urban and transportation models (Wilson, 1970).

Constructing this framework, we have borrowed much from existing well-established techniques and models from statistical physics and have striven to retain the original terminology as far as possible because we feel that this will promote clarity in the long run. There is, however, because of this, some need to introduce certain concepts in this new context. In particular, we want to clarify our use of the term 'interaction' because there is no complete agreement between the meaning of that term in all fields and we were unable to find a good synonym: as defined in physics, interactions are the influence (often in terms of force) that entities have on each other. The intensity of an interaction is often a function of the distance between the interacting parts. In the context of an urban system, it would correspond rather to a much simplified aggregate of several different phenomena that in various ways act to impose influence between land uses in different areas (for example: competition between businesses, noise and pollution from industry, employment, etc). The term 'interaction', as used here, does not specifically refer to any economic or demographic measurement of exchange between different areas.

This integration of theory and realism is crucial because, in general, more powerful tests of theory demand more detailed—that is, realistic—predictions. Conversely, realistic models developed for practical applications by planners and others can only be used with confidence if they are based on well-tested theory and thus come with some assurance of reliability. Furthermore, some long-standing theoretical issues in urban spatial dynamics have practical policy implications. For example, the well-known rank–size rule for urban size hierarchies, where the relationship between the size of a center and its rank is log–linear, still lacks a complete theoretical explanation; and those concerned with policy consider it uninteresting and arbitrary. Yet some major planning and policy initiatives have in effect been directed at changing or preserving the rank–size relationship in particular systems: the attempt in France, for example, to reduce the dominance of Paris (that is, to nudge the urban system toward the log–linear relationship) and the policies in many European and North American cities designed to preserve a strong central business district (that is, to prevent the collapse of the log–linear relationship among intraurban retail centers) (White and Engelen, 1997).

Recent work in urban systems theory, in particular, and in the area of complex systems, in general, is providing deeper theoretical insight into the origin and evolution of urban form. Results in the area of complex systems suggest that the rank–size relationship may be a characteristic scale-free metastable form transitional between clustered and dispersed (Kauffman, 1993; Langton, 1992). In this paper we continue this line of investigation, and using the urban model presented here we explore the origin of urban structure by treating it as the result of different types of fundamental dynamics and the phase transitions between these different dynamics.

Phase transitions between different universal classes of behavior are brought about by varying what corresponds to the 'temperature' and 'pressure' of land-use interactions within the system. We first investigate a simple situation with a single growing land-use class up to more complex setups with multiple classes. These different behaviors of the system can be viewed as different physical states of aggregation.

The model we propose is lattice or raster based, and has much in common with a standard CA. The lattice dynamics is interaction based, and takes into account interactions that decay exponentially with distance over the entire lattice and uses a modified Markov random field (MRF) approach for the dynamics (Clark, 1951). The land-use potentials for change are calculated locally, but the actual land-use changes are randomly selected globally. The potentials are calculated from pairwise first-principle interactions between cells. This approach builds on the update rule used by White et al (1997) where interactions are to be viewed as an aggregated net effect of many various factors where human decisions and behavior are ultimately defining.

The addition of an edge-growth rule allows us to produce structures similar to those generated by DLA models. Allowing a balance between edge growth and emergence of urbanization away from already urbanized areas defines a model very similar to that of Clarke et al (1997). The edge-growth rule requires, or at least very strongly favors, the addition of urbanized cells to take place only adjacent to other urban cells.

2 Simulation framework

Each part of the simulation arguably corresponds to elements of the real-world system of which we are representing the dynamics. In each of the following sections we shall argue for the basic structure of our framework.

The simulation technique used borrows much of its theoretical background from statistical physics and complex systems. Because the theoretical foundation of the model largely belongs in these fields, we have in most cases retained the original terminology. Some of the terms used are susceptible to misunderstanding because they carry different meaning in different fields. Throughout, terms such as temperature, density, pressure, and energy are to be interpreted as pertaining to the low-level workings of the urban system (as defined here) and not to the energy use or current temperature of the city. The analogies have nothing to do with refined energy such as gasoline or electricity usage.

The aim of this work is to produce a simulation framework that integrates the levels of detail that actual scenario prediction demands while still being well-defined mathematically. Thus, our use of these terms does not exclude or contradict other meanings of the same words in other fields of research.

2.1 Basic dynamics

Formally, our simulation framework consists of a modified 2D MRF (stochastic cellular automaton) representation of the site–site interactions using a recursive mean-field approach to take into account interactions not only from neighboring sites, but from all lattice sites. Averaging is a standard method in statistical physics and has also been explored in urban models (for example, see Batty, 1976; Broadbent, 1971). The state transitions at the individual sites are determined by a global (probabilistic) selection criterion, as in evolutionary selection, and not by a local selection criterion as in the classical MRF.

The regional settlement dynamics in each iteration of the simulation can be summarized as:

$$S(t+1) \;=\; S(t) - \Delta R(t') + \Delta A(t''), \tag{1}$$

where $S(t)$ is the global system state (lattice configuration) at time t, $\Delta R(t')$ the activity removal dynamics, and $\Delta A(t'')$ the activity addition dynamics at times t, t', and t'', respectively, with $t < t' < t'' < t+1$. Obviously, this defines a traditional evolutionary dynamics with removal of the least fit activities followed by the addition of the most fit activities within the population of available cells.

2.2 Lattice and land use

We represent land as a 2D square grid divided into N cells which are equally sized square patches of land. In principle, any area can be simulated, although we have used only square lattices here. Each cell corresponds to an area and its (discrete) state represents what is on it—its land use. This means that we divide all possible uses of land into classes which are assumed to be homogeneous within a cell and to interact with each other in a homogeneous manner. In the simplest case, only two classes are considered—built and rural—but this setup can accommodate any resolution of land use because there is no limit to how many states a cell can take. Besides a simple two-state land-use representation we also discuss more complex setups with multiple land-use classes.

2.3 Markov random fields

A classical MRF (Kindermann and Snell, 1980) representation of a 2D land-use dynamics may be defined as follows: from a set of c land uses, consider two different land uses $a, b \in \{1, 2, \ldots, c\} = C$. The maximum radius of land use to land-use influence is R. The potential ('energy') of land-use class a at a given location x is

$$E_a(x) = \sum_{d \leqslant R} \sum_{b \in C} w_{ab}(d), \tag{2}$$

where $w_{ab}(d)$ is the positive or negative influence ('energy contribution') from land-use class b to land-use class a (including transportation infrastructure) at distance d.

In order to be able to compare the suitability of cells we want to map them to the unit interval. One way of doing this is by using the Maxwell–Boltzmann transformation. However, we use a positive exponent to preserve the analogy with a measure of goodness. This means that we translate negative potential energies into low probabilities and positive energies into high probabilities. To express such land-use transition probabilities—say from land use b to land use a—one can use

$$p_a = \frac{\mathrm{F}[E_a(x)]}{\sum_{b \in C} \mathrm{F}[E_b(x)]}, \tag{3}$$

where F is a Boltzmann transformation

$$\mathrm{F}[E_a(x)] = \exp[\beta E_a(x)], \tag{4}$$

and β is a free parameter which corresponds to T^{-1}, where T is the 'temperature' of the system. Equations (2), (3), and (4) define a Markov chain in each lattice point. See figure 1(a).

In the urban setting the temperature corresponds to the degree to which a given activity at a given land location takes into account the influence of neighboring activities. If the affinity between two land-use classes is thought of as a statistical mean, the temperature acts much like the standard deviation of the distribution.

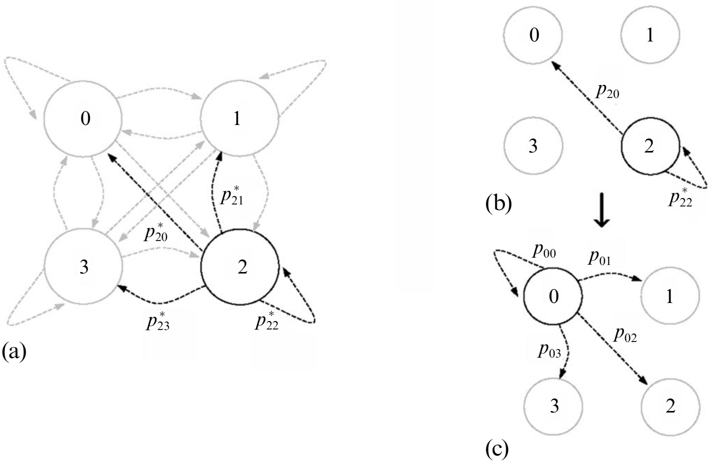


Figure 1. The land-use transition mechanism that is used in the model can be formalized as a classical Markov random field with Markov chains as shown in (a). However, the removal and the addition processes are modeled separately and the update cycle that corresponds to one transition in (a) is split into two phases: removal (b) and addition (c). As an example, we here consider a transition between land-use class 2 and land-use class 3. During the update cycle we first perform the death process (b) (see section 2.6.1 in the text) in which developed cells are considered for change into the undeveloped state. During the second phase (c) (section 2.6.2) the undeveloped cells may be changed into developed states. Taken together, they form a traditional Markov chain, as shown in (a), where the transition probabilities are defined from those in the two-phase update cycle as $p^*_{ab} = p_{a0} p_{0b}$.

The higher the temperature the less influence the neighborhood has on the location and eventually, for high temperatures, the dynamics turn into a pure random mapping where all transitions are equally probable. The β in the Maxwell – Boltzmann distribution stems from a Lagrange parameter in the calculation of a model of minimal assumption (or maximum entropy). An interpretation of this parameter in information theory as a measure of lack of information is consistent both with the interpretation from statistical physics and with one that makes sense in the context of urban dynamics. If the temperature is high, the system is only weakly determined by the model and a low temperature means the converse.

Although the fundamental idea behind our approach is as described above, it differs in three significant ways: (1) we take the entire geographic region into account rather than just a small radius neighborhood, (2) the state transition probabilities are defined globally and not locally, and (3) we split the activity change dynamics into two processes: activity removal and activity addition as expressed in equation (1).

2.4 Mean fields and long-range interactions

Why take into account long-range interactions and why make mean-field approximations (averages) of the distant interactions? First, we believe that long-range interactions are significant because many of our most important activities may not be located in the immediate neighborhood of where we live. Our workplace, where we go shopping, educational institutions, and where we spend recreational time can be far away from where we live and still influence where we decide to settle. Second, we believe that the further away from a given location a particular activity is located, the weaker is its influence on the original reference location. Also the precise location of a particular activity becomes less important the further away it is from the reference point. These assumptions make it possible to define simple long-range land-use – land-use interactions. A clear simplification is that we use an Euclidean geometry. In reality, distance is more analogous to travel time so the transportation network distorts the geometry.

In order to achieve a fast update algorithm when taking into account long-range interactions, we use a recursive set of mean-field approximations (that is, spatial averages) of distant areas rather than taking each cell into account individually. This is done through a successive definition of aggregated lattices at different levels.

The fundamental lattice with N lattice points is referred to as the level-0 grid, which is the grid with the highest resolution. Grids at higher levels, l, of aggregation have cells that are mean fields (averages) of progressively larger concentric portions of the level-0 grid. Thus, an l-level cell is contributed to by 3^2 times as many level-0 cells as an $(l-1)$-level cell. These recursive levels are defined in figure 2. Starting from the most coarse grained, or aggregated, level, L, where the whole lattice is aggregated, 3^2 new subgrids are generated for each recursion and thus

$$N = (3^2)^L \Leftrightarrow L = \frac{1}{2}\log_3 N, \tag{5}$$

which indicates that $\frac{1}{2}\log_3 N$ recursive lattice-averaging operations are needed for the update of each site. L then defines the depth of the lattice.

The cell count of activity a at location (x) on the atomic level is binary,

$$c_a^{(0)}(x) = \begin{cases} 1, & \text{iff land use} = a, \\ 0, & \text{else}. \end{cases} \tag{6}$$

From this we then define mean fields $\langle c \rangle_{a,l,x}$ as the count of cells of land-use class a in a square region with side 3^l centered at x.

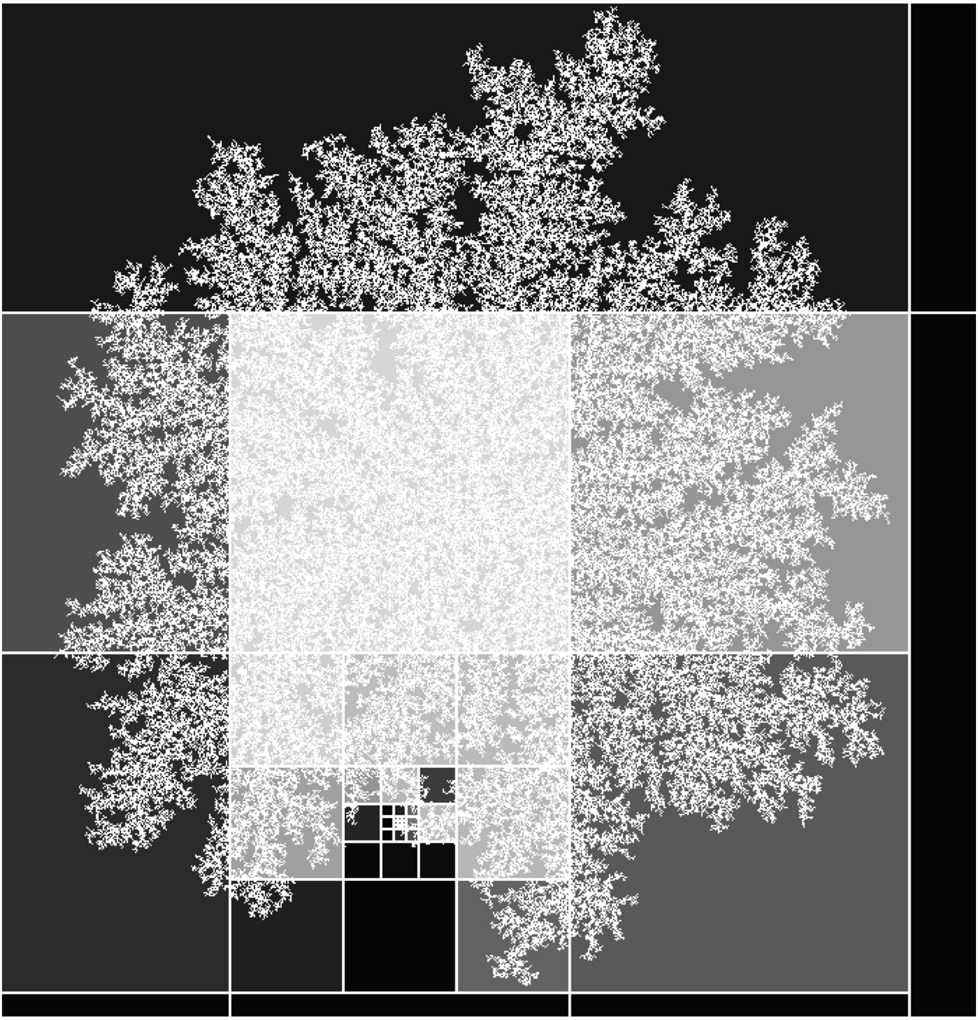


Figure 2. An illustration of the multilevel mean-field representation is here made showing the detail with which the center cell perceives its surroundings. The grey fields that overlay the configuration in the background correspond to mean fields, $\langle c \rangle_{a,l,x}$, where a is the single land use in this configuration (developed) and x is the center of the field. Distance averaging is realized by treating all cells within a mean field as being located at the center point x of that field.

We have made the simple assumption that we perceive our surroundings in successive scales of detail, paying more attention to the specifics of what is close by than to what is far away. The conception of using scales is appealing for many reasons, not least because fractal structures seem to be inherent in many natural phenomena. We can also trace this reasoning by looking at human concepts of geographical/economical/political areas: continents, nations, regions, cities, suburbs, blocks, etc are all conceptual levels of scale at which we view our world. Indeed, the idea of a hierarchy of successive scales is well known and has for long had an important role in geography through central place theory (Christaller, 1933).

Based on this assumption, the algorithm we use for updating the grid has a time complexity in $\Theta(N \log N)$ rather than in $\Theta(N^2)$ for a naive algorithm that would let all cells interact with each other individually.[(1)] The reason for this improvement is that, instead of relating all cells to each other (N^2), we recursively aggregate mean fields with a magnification factor of 9 (3^2) at each level. Grids are of sizes $N = 3^{2L}$, so we can see that mean fields the size of the whole lattice will be aggregated in L recursions. Looking at equation (5) we see that $L = \frac{1}{2}\log_3 N$. Thus, performing this operation for each of the N cells of the lattice gives us $N\frac{1}{2}\log_3 N \in \Theta(N \log N)$.

2.5 Activity interactions

As we shall see, updates are based on fields that are calculated from interactions between cells on the grid. The interaction strength between a pair of land-use classes is a function of the distance between the interacting cells. Land-use classes can be either dynamical or static. Static land uses affect dynamical ones but do not change themselves. Cells belonging to a dynamical land-use class both change and affect other cells. Examples of static land uses are structures such as railroads, or highways. Examples of dynamical land uses are housing, industry, or commercial.

To make the interaction function more manageable, it is divided into two functions. One models how the importance of the individual activities decays with distance as more and more activities exist for each distance class. This exponential decay, $d^{(l)}$, is common for all land-use classes and is defined as

$$d^{(l)} = 3^{-\gamma 2l}, \tag{7}$$

which exactly accounts for the exponential increase in cells at each recursive level of interaction with a free adjustment parameter $\gamma = 1$. The influence of γ is shown in figure 3. As can be seen in equation (7) the influence of one cell on another tapers off exponentially with distance, depending on the parameter γ. This exponential decay turns up in population density (Clark, 1951) and is taken for an empirical fact.

The other part of the interaction function models how one particular activity influences a particular other activity at different distances when all other activities are ignored. We term this 'affinity', as it defines the microscopic pairwise (cell-to-cell) interactions.

Throughout the paper we will use the following brief form for representing affinities:

$$A_{ij} = \left\{ A_{ij}^{(0)}, A_{ij}^{(l)}, \dots A_{ij}^{(v)} \right\}, \qquad v = \frac{1}{2}\log_3 N, \tag{8}$$

where $A_{ij}^{(l)} \in A_{ij}$ is the attraction that land use j exerts on land use i at scale l. For each level of aggregation l and each pair of activities a and b we define an affinity as $A_{ab}^{(l)}$. Positive function values of the affinity function mean that cells of land use b stimulate the growth of land use a at a distance of 3^l. Conversely, negative function values inhibit growth. However, not all of these affinities have been measured directly and thus they have to be approximated in some manner. In simulations that have been applied to real-life problems, parameters have been tuned by using the quality with which the model recreates the present with the past as a utility function. When good performance is achieved, the parameters may be such that the simulation is able to generalize the behavior of the desired system and to extrapolate the urban form and structure into the future with some accuracy (Clarke et al, 1997).

The resulting interaction strength as a function of land-use pair and distance is then defined by

$$I_{ab}^{(l)} = d^{(l)} A_{ab}^{(l)}, \tag{9}$$

which corresponds to w_{ba} in equation (2).

[(1)] Technically $O(N \log N)$ would also be correct, although $\Theta(N \log N)$ is more accurate.

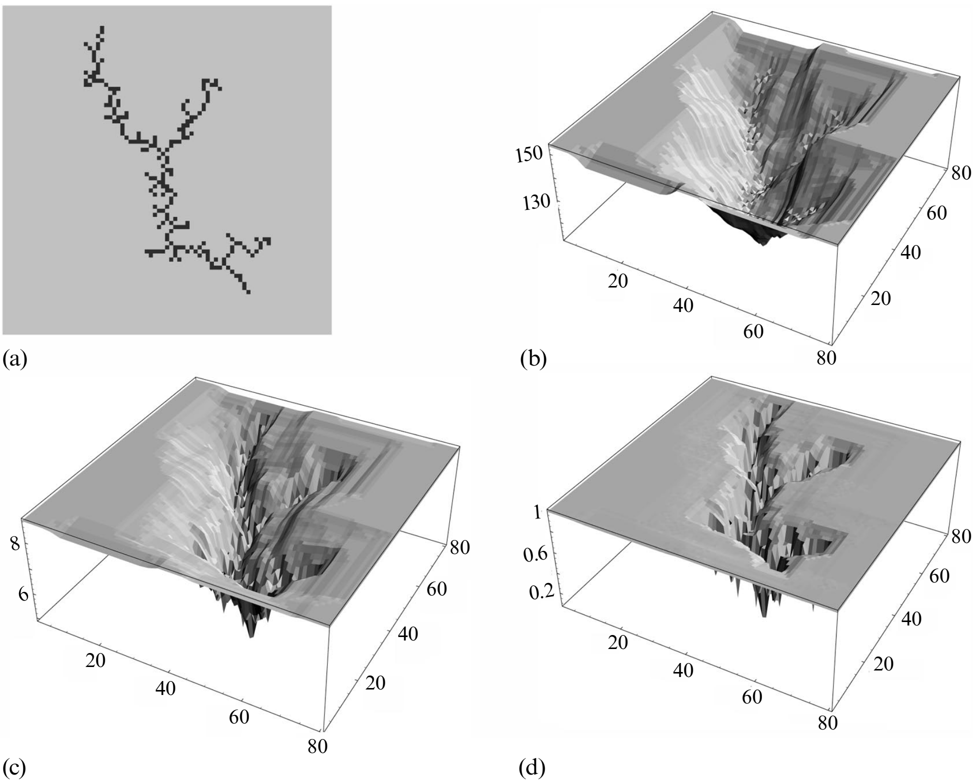


Figure 3. Provided that interaction lengths are constant, tuning γ lets us weigh the importance of each distance class. The greater the value of γ the shorter the effective interaction radius. Shown is a series of energy landscapes all calculated from the lattice configuration shown in (a) with values of (b) $\gamma = 0.25$, (c) $\gamma = 0.75$, and (d) $\gamma = 1.25$. The urban cells are inhibited by each other rather than attracted under the 'unwilling neighbors' rule; hence the canyon-like landscape (see section 4.1 in the text). Note that the concept of energy adopted here corresponds most closely to land-use potentials in the tradition of urban dynamics research and the concept of Darwinian fitness. Thus, rather than seeking energy minima, urban cells seek to maximize the 'fitness' of the cells they occupy.

2.6 Evolutionary selection

For each cell and activity a a corresponding potential (energy) $E_a(x)$ is calculated using the recursive mean-field scheme:

$$E_a(x) = \sum_{l=0}^{L} \sum_{b \in C} I_{ab}^{(l)} \langle c \rangle_{b,l,x} , \tag{10}$$

where $I_{aa}^{(0)}$ is zero, and $I_{ba}^{(0)}$ expresses the inertia or potential energy cost associated with an activity change from a to b. We assume the same cost for all activity changes. Equation (10) gives a measure of how desirable each cell is for each dynamical land use at this particular time. Again, we need to transform these potentials into probabilities and using the Boltzmann transformation gives us a simple mapping with the desired properties. Recall equation (2):

$$b_a(x) = \exp[\beta E_a(x)] . \tag{11}$$

Recall equation (4).

The weights $b_a(x)$ are modified by an aggregated suitability measure $\alpha_a(x)$ which includes layers of terrain, zoning, slope, and other factors:

$$b_a(x) = \alpha_a(x) b_a(x) . \tag{12}$$

The weights $b_a(x)$ may also be modified by regional constraints given by such demographic and economic factors as the total population and available workforce. We assume such constraints to be externally determined by scenarios.

To simplify the simulation we can naturally separate the dynamics into two processes: (1) activity removal and (2) activity addition. These two processes of birth and death are of course closely coupled as, for example, in an activity change at a particular location which in this context consists of an activity removal followed by an activity addition in the same update. Removals leave vacant land and activity addition can only occur on vacant (nonurbanized) land. The activity removal and the activity addition processes differ in one important aspect. The activity addition consists mainly of a global selection of the most suitable site among many possible in the urban area. The activity removal stems mostly from a local determination of the desirability of the current activity compared with the alternatives. Two extremes exist where either all activities are nondesirable or where all activities are highly desirable. In either case the relative imbalance between the different activities determines the outcome as these local imbalances are compared globally.

2.6.1 *Death process—activity removal*

Because of the importance of local imbalance between the desirability of current activity on a given site, the activity removal process is defined through two steps. The first step

$$s_a^R(x) = \frac{\exp[\beta E_a(x)]}{z(x)} , \tag{13}$$

where

$$z(x) = \sum_{b \in C} \exp[\beta E_b(x)] , \tag{14}$$

defines the local desirabilities (probabilities) where the most desirable activity has the highest value. As a second step the transformation

$$\left[s_a^R(x)\right]^{-1} = \frac{z(x)}{\exp[\beta E_a(x)]} = \sum_{b \in C} \exp\left\{\beta[E_b(x) - E_a(x)]\right\} \tag{15}$$

defines the least desirable activity with the highest value. By normalizing the local reciprocal activity desirabilities over the entire region we obtain global probabilities similar to equation (13) which expresses the 'fitness landscape' for activity removal. The local activity-removal fitness is now

$$s_a^{R*}(x) = \frac{\left[s_a^R(x)\right]^{-1}}{Z^R} , \tag{16}$$

where

$$Z^R = \sum_{b \in C} \left[s_b^R(x)\right]^{-1} . \tag{17}$$

In section 2.3 transition probabilities from activity a to undeveloped land are defined for the MRFs. Computationally, the activity removal process is defined in the following manner. We need to map a random number to a cell with a probability given by equation (13). Then

$$\Sigma_a^R = \sum_x s_a^{R*}(x) \tag{18}$$

defines the sum of all the Boltzmann-transformed potentials for a certain land-use class a. We now store all partial sums of a summation over the whole lattice

$$\sigma_a^R(k) = \sum_{x=1}^{k} s_a^{R*}(x) . \tag{19}$$

Then we select a uniformly distributed random number $r \in [0, \Sigma_a^R)$ and investigate for which k, $\sigma_a^R(k-1) \leqslant r < \sigma_a^R(k)$. The index k then gives the cell we change. See appendix A for an example.

In the same way as for the activity addition process [equation (24) below], to ensure reasonable competition for activity removal, we switch appropriately between the different land-use classes in the selection process to generate undeveloped sites.

2.6.2 *Birth process—activity addition*

Conceptually, the activity addition dynamics is defined through a global selection of cells based on their 'fitness'. For each land-use activity a

$$s_a^A(x) = \frac{\exp[\beta E_a(x)]}{Z_a^A} \tag{20}$$

defines the local fitness normalized over the entire region

$$Z_a^A = \sum_x \exp[\beta E_a(x)] . \tag{21}$$

Note that $s_a^A(x)$ defines the transition probabilities to activity a for the MRF over the vacant or undeveloped cells.

Analogous to the activity removal process the activity addition process can be computationally defined through

$$\Sigma_a^A = \sum_x s_a^A(x) , \tag{22}$$

the global sum of all local probabilities for a certain land use a together with the stored partial sums of these same local probabilities summed over the whole lattice

$$\sigma_a^A(k) = \sum_{x=1}^{k} s_a^A(x). \tag{23}$$

We select a uniformly distributed random number $r \in [0, \Sigma_a^A)$ and investigate for which k, $\sigma_a^A(k-1) \leqslant r < \sigma_a^A(k)$. The index k then gives the cell to which we add an activity.

To ensure an appropriate site competition between the different activities in the multiactivity simulation a switch between their associated selection processes is defined through

$$P_a^A(t) = \frac{U_a(t)}{\sum_{i=0}^{|U|} U_i(t)} , \tag{24}$$

where $U_a(t)$ defines the (here externally determined) temporal changes in land use a. As equation (24) is updated after each selection, the correct number of activity additions can be ensured. Because the regional growth or contraction is determined by the number of additional activities minus the number of activity removals we still need to define (in section 2.7) how each of these processes can be derived from the overall activity changes.

The following example clarifies the importance of randomizing the order of addition as well as removal: assume that, in a simulation, land-use class 'residential' happens to have an index lower than that of 'commerce'. If we were consistently to add the cells of the residential land-use class before those of the commercial, the lattice points that happen to be in demand by both will more often be occupied by residential. Any correlated order we choose will result in bias. The importance of randomizing the order of changes during an interval when a discrete representation of time is used is a common problem, not specific to urban systems.

2.7 Settlement dynamics

The regional settlement dynamics in each iteration of the simulation are summarized in equation (1). Obviously, this defines a traditional evolutionary dynamics with removal of the least fit activities followed by the addition of the most fit activities within the population of available cells. Both the activity removal and the activity addition are characterized by random sequential processes which are running without a recalculation of the global systems state. Thus, this simulation update is similar to an Euler update of a differential equation and as such only small changes to the system state can occur in each update. Technically, this means that at most a few percent of the activities can be changed in each update. In the simulation where this applies, we have used 1–10% global changes in each update. The relative difference between the activity addition and removal process defines the urban growth or contraction and the fraction of cells with activity changes determines the land-use turnover. Note that the same growth can be obtained through many different addition and removal combinations as long as the difference between the activity removal and addition is constant.

3 Urban phase transitions

To tackle the multitude of detailed behaviors of a dynamical system, typical properties of the dynamics, the so-called 'universal classes of behavior', are often sought. The question we ask when doing so is: "Are there significant portions of the parameter space throughout which the system seems to behave in a qualitatively similar way?" Maybe there is a measurable property of clustering or some other statistical property that can be observed and used as an order parameter. When the system undergoes a transition from one phase to another, it suddenly starts to behave in a fundamentally different way. Ideally, as defined for an infinite and steady-state simulation, the transition in fact constitutes a discontinuity in some order derivative and thus there is indeed no interval between the two behaviors. Despite not being strictly defined for these simulations, because cities are nonequilibrium, growing, structures, the basic feature where large parts of the parameter space exhibit a uniform behavior is still clearly present.

3.1 Urban phase transitions

For a dynamical system with as many states and state interactions as our urban settlement simulation we should expect the existence of multiple types of dynamics (universality classes) possibly separated by phase transitions. And this is indeed true for our simulations.

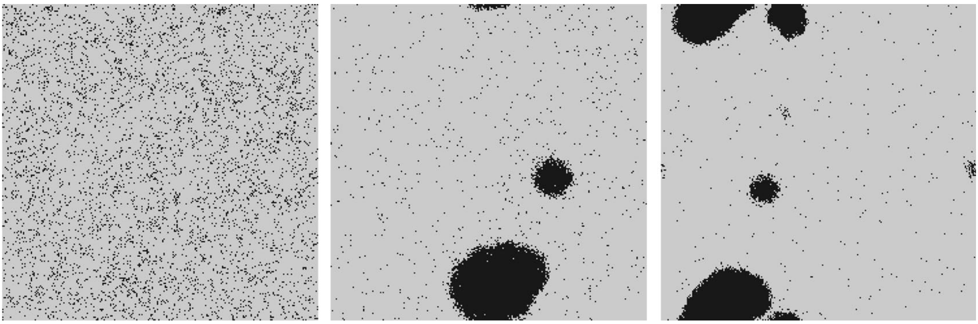
Perhaps the best known and the simplest classification of different types of dynamics is the distinction between matter's three phases: gas, liquid, and solid. Each of these phases is characterized as belonging to different universal classes of dynamics because they behave in fundamentally different ways and are separated by phase transitions. Other well-known physical examples of phase transitions can be found within liquid mixtures such as water and oil. Oil and water will spontaneously separate into two phases, but if detergent (soap) is added this separation is organized as a mixture of small oil droplets with detergent at the oil–water interface, dependent on the amount and type of detergent. Analogs of all the above transitions found in simple physical systems can easily be located within the urban settlement dynamics as we have defined them. As we discuss such transitions between these qualitatively different types of urban settlement dynamics we also discuss what they mean in the urban setting.

It should be noted that phase transitions are not easily defined for nonequilibrium dynamical systems such as a growing city. Still, the mechanism we are looking at is the same, and it is evident how the tuning of, for example, the 'temperature' will take us through sharp transitions between two different universal classes of behavior in the urban settlement dynamics.

Idealized simulations with a uniform background as we use here are too simplistic to produce cities that look much like what we might expect to find on a map. However, the purpose of this exercise is to discuss (1) some of the rich dynamics this type of simulation can generate, (2) the transitions between these qualitatively different types of dynamics, and (3) what it means for real urban dynamics.

3.2 Settlement temperature

For the simple, single land-use class simulations we use a generalized affinity function defined as $A_{00} = \{0, 1, 1, 1, 1\}$ [equation (8)]. As we vary β or T^{-1} as defined in equation (4) and keep the number of urbanized cells constant over time we observe a significant shift in all observables as reflected in figure 4 and color plate 1(a).



(a) (b) (c)

Figure 4. Temperature phase transition: (a) gas, (b) near-transition, and (c) liquid. Spatial correlation is plotted against distance and β (T^{-1}) on a 243×243 lattice.

This dynamic indicates a first-order transition from gas to liquid as the gas condenses into droplets which merge and eventually form a single aggregate. As the temperature decreases further, the condensation process is faster, and more metastable local clusters appear. The lattice size used in these simulations (figure 4 and color plate 1) is $(3^5)^2 = 243 \times 243$ and the simulation time is very long (3000 updates) to ensure a steady state. In each update 10% of the urbanized cells are redistributed [$\Delta R = \Delta A$ in equation (1)].

Because of the steady-state condition these transitions do not map directly onto a realistic urban growth situation because real urban systems presumably never reach a steady state. Real city growth should be considered to exhibit nonequilibrium, transient dynamics. The difference between a steady-state condition and the transient dynamics is partly illustrated as we compare figure 5 with color plate 1 (see over).

In figure 5 (see over), a constant number of randomly located urbanized cells are used as initial conditions. As can be seen from the figure, the cluster formation occurs

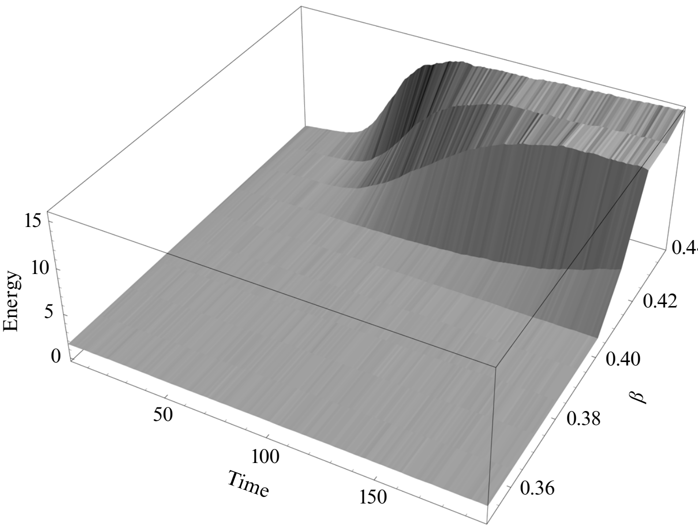


Figure 5. Dynamics of a temperature transition: energy plotted against time and β (T^{-1}) on a 243×243 ($243 = 3^5$) lattice. Note how the transition between the two states of the system (liquid and gas) occurs earlier and gets sharper the further the temperature drops below the transition temperature (which corresponds to β about 0.38).

very quickly for high β values (low temperatures) and it does not occur for values below the phase transition value $\beta \sim 0.38$.

The temperature in the urban settlement dynamics corresponds to the sensitivity with which a given land-use type cares about its neighborhood. The lower the temperature the more attention is paid to the local land-use environment. For temperatures above criticality, no attention is paid to the local land-use structure, and the resulting land-use dynamics becomes a random mapping between the possible land uses. Obviously, all known cities are presumably below this transition although small, local areas within a city may be thought of as having a high temperature in the modeling sense.

The history of human settlement holds many examples of what can be thought of as temperature transitions. When large (ancient) cites disintegrated they did so because the mutual advantage of, or even the possibility of, living in a cluster disappeared. Another way of thinking about a high-temperature situation regionally could be an idealized urban area where transportation time does not factor in. Modern information technology (IT) such as telephones and computer networks are excellent examples of recent technology that diminishes the importance of distance and thereby possibly increases the temperature of the system under our definition. Indeed, companies that provide services mainly via IT, such as computer support providers, often operate from rural areas where the workforce is cheaper.

3.3 Settlement pressure (density)

For a constant increase in the total number of urbanized cells on a given lattice, which defines the pressure, we again observe what looks like a first-order transition from a gas phase to a liquid phase as the urbanized cells start to aggregate. See color plate 1(b). The autocorrelation function is fairly constant below the critical pressure (about 4850 cells on a 243×243 lattice) and as clusters begin to form it rises and eventually levels off as clusters are formed (above 5200 urbanized cells).

The pressure transition from gas to liquid can be observed in many contemporary urban growth situations. Such a transition can, for instance, be seen along the city edge when two urban areas grow into each other. At a certain point the landscape between the formerly separated developed areas fills in on a much shorter timescale than that required for the two urban areas to touch each other. This has, for instance, been seen in the cities of the Ruhr district in Germany after the Second World War. This can also be observed at a much smaller scale as 'infill' between 'arms' of modern urban sprawl in the southwestern part of the USA.

3.4 Multiple land-use interaction transitions

Yet other types of phase transitions can be seen for multiple land-use dynamics. We now discuss a sequence of transitions which are analogous to samples of a phase diagram for three different fluids. As 'control parameters' we do not use temperature or pressure, but the interaction functions between the different land uses. For the sake of discussion assume that these three land uses represent housing, commerce, and industry. See color plate 2. Initially they all mix well which can be mapped onto similar, mutually attracting affinities, which means that we have urban clusters with mixed land uses [color plate 2(a)].

If we now change the housing–industry interaction such that housing dislikes industry, pure industry clusters emerge, together with other clusters with well mixed commerce and housing. This situation is analogous to the phase separation between oil and water [color plate 2(b)].

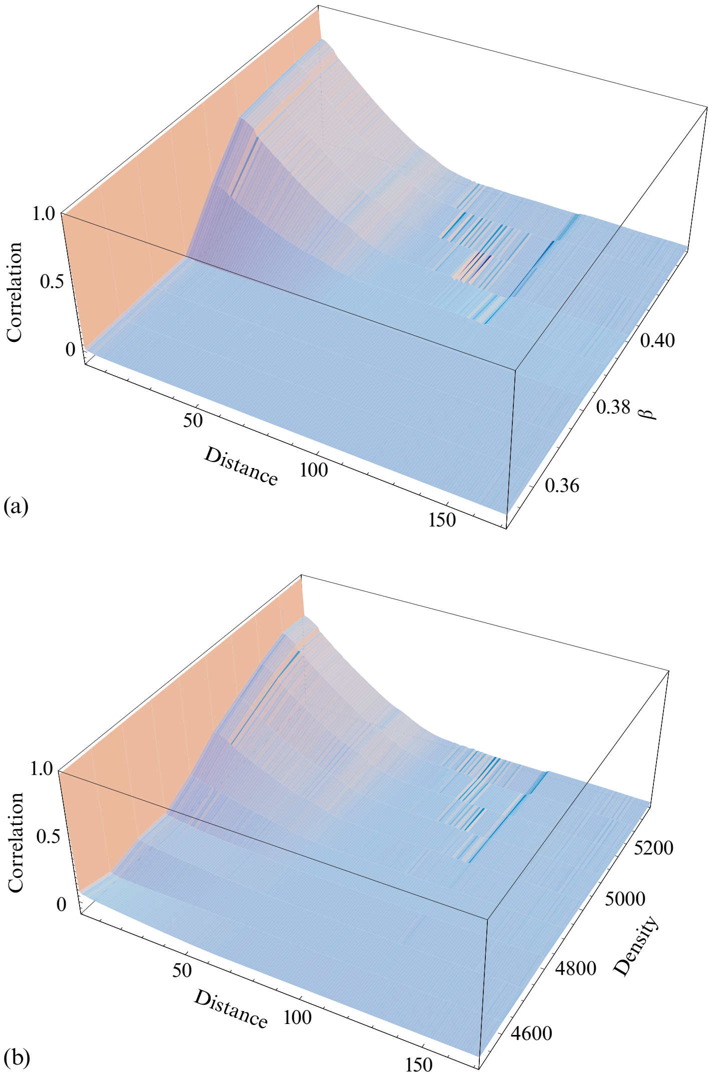
Changing housing to dislike commerce nearby, and changing commerce to dislike commerce far away, we obtain a situation where the three land uses can coexist in several different ways. Here we observe structured clusters with all land uses present where housing and commerce are neighbors which again are neighbors with industry. This situation is analogous to oil in water with a surfactant at the interface between the two. The dynamics also allow clusters with only housing and commerce and clusters with only industry and commerce. Note how the commerce can be completely 'dissolved' within the industry cluster and how the commerce is clustered ('nonsoluble') within the housing cluster [color plate 2(c)].

Transitions as generated here are observed in many places in the world. In many ways an artificial separation between land uses is forced on the urban dynamics through zoning regulations. Mixed-use areas are seen in many older cities in Europe and Asia whereas a clean separation of land use is more recent. It is nevertheless represented on all continents as, for example, suburban growth and the downtown business district.

4 Modifications of the basic Markov random field dynamics

Using a simple additional 'stickyness' rule under which new urbanized cells are constrained to appear only next to other urbanized cells we can generate structures that are very similar to those resulting from a diffusion-limited aggregation (DLA) process (Batty and Langley, 1994; Witten and Sander, 1983).

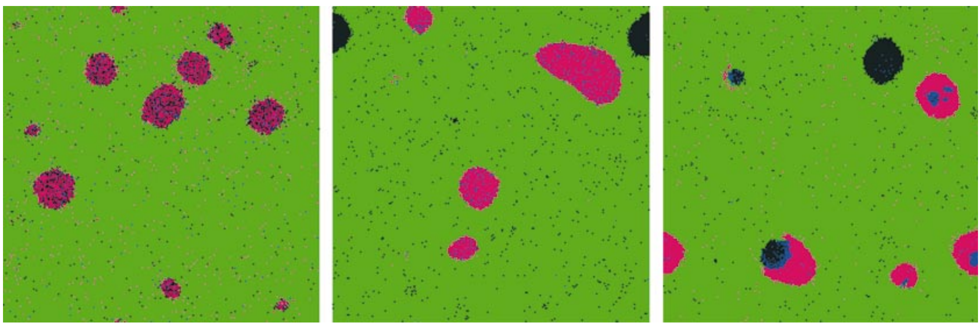
If this edge-growth rule is complemented by a small additional chance for urbanization away from any urbanized area we define a model very similar to the urban growth model developed by Clarke et al (1997). Formally we obtain this dynamics from the simple MRF formulation by introducing a parameter ϵ, $\epsilon \in [0, \frac{1}{2}]$. We multiply the calculated MRF edge probabilities for each land-use class with $(1-\epsilon)$ and the other MRF potentials away from urbanized cells with ϵ, where ϵ is typically very small.



(a)

(b)

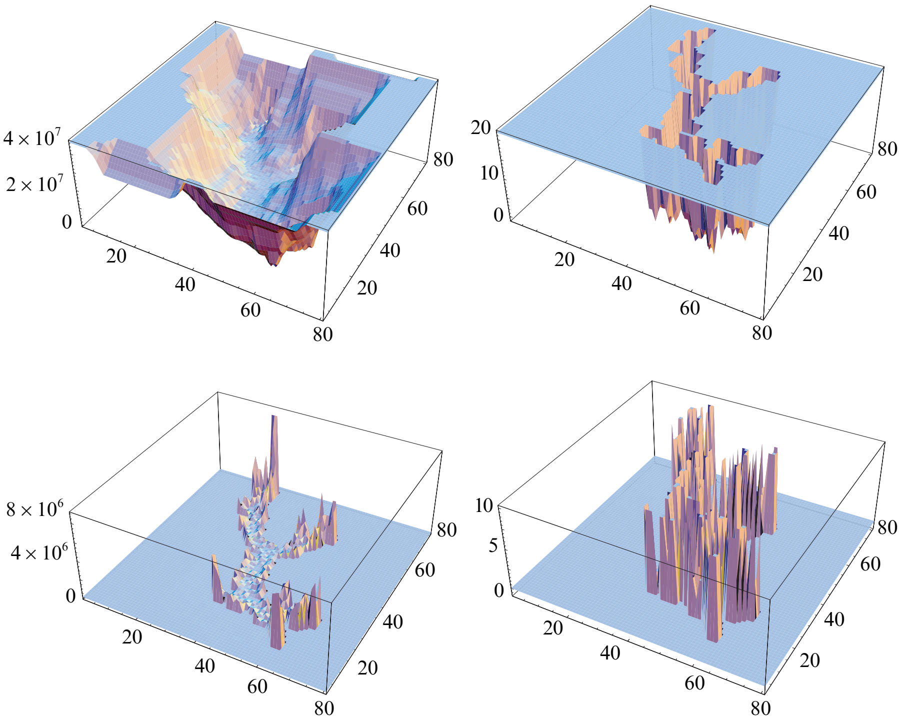
Color plate 1.



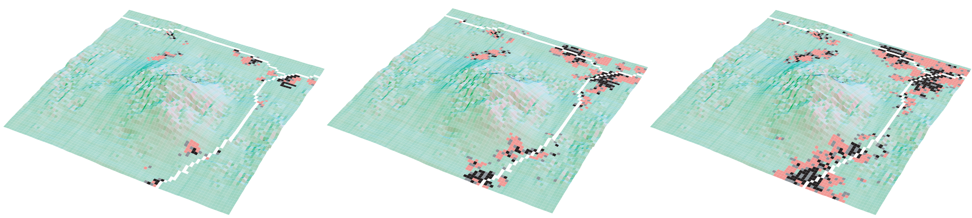
(a) (b) (c)

Color plate 2. Black, blue, red, and green cells are industry, commerce, residential, and undeveloped, respectively. Transitions between different universal classes of behavior are here brought about by changing the affinity component of the interaction functions [Equation (8) in the text]. In configuration (a) all land-use classes stimulate growth of all land-use classes. In (b) a transition where industry is clustered separately is brought about by altering one of the affinity vectors correspondingly. In (c) two additional affinity vector changes give rise to a fairly complex spatiotemporal behavior where clusters move over the lattice and instabilities cause sudden shifts: residential dislikes commerce near and commerce dislikes commerce far.

Color plate 1 (see opposite). Autocorrelation plot (association of land use with itself) for single land-use phase transitions (243×243 lattice and steady-state conditions with equal amount of addition and removal): (a) β (temperature) sweep; (b) density (pressure) sweep. The correlation measure provides a picture of the difference between the observed distribution and a background distribution, which is taken to be an entirely random pattern. A positive correlation (z-axis) for a certain distance (x-axis) means that at that distance from a developed cell the likelihood that an arbitrary cell is also developed is higher than for a random pattern. The converse is true for negative correlations. In the plots we can see that for low values of β and total developed area, the configuration is found to correspond well to a random configuration (correlation ≈ 0 for all distances). Then, as β or the number of developed cells increases, the short-scale correlation jumps to a higher value. This indicates a transition from nonclustered to clustered patterns.



Color plate 3. Potential landscapes calculated from figure 8(a) are shown. By landscapes we mean that each lattice point in the figure corresponds to a cell in a configuration. The figures on the left show landscapes calculated using all range classes and those on the right show landscapes using only close-range interactions. The top landscapes show nonnormalized probabilities for site selection before application of the $(1 - \epsilon)$ factor and the bottom images depict the landscape after the application of that rule. Because urbanized cells repel further urbanization under the 'unwilling neighbors' rule, the inclusion of long-range interactions will cause interior parts to receive more negative contributions compared with exposed parts. Looking at the bottom-left figure, it is evident how protruding parts of the configuration are more prone to growth than are interior, more shielded, parts.



Color plate 4. In this time-series we see the growth of 'Rockville'. The topography is computer generated and suitabilities [equation (12) in the text] are derived for each site as a function of slope. The cells correspond to land areas with a side of roughly 100 m, the simulated growth has taken place on a time scale of a few decades. The configuration is seeded only by the two limited-access roads that run north – south and east – west. Note that the vertical relief of these images is exaggerated by a factor of two.

Note that at its limits, ϵ corresponds to only edge growth and no preference for edges, respectively. Recalling equation (20), we can formulate this transformation as

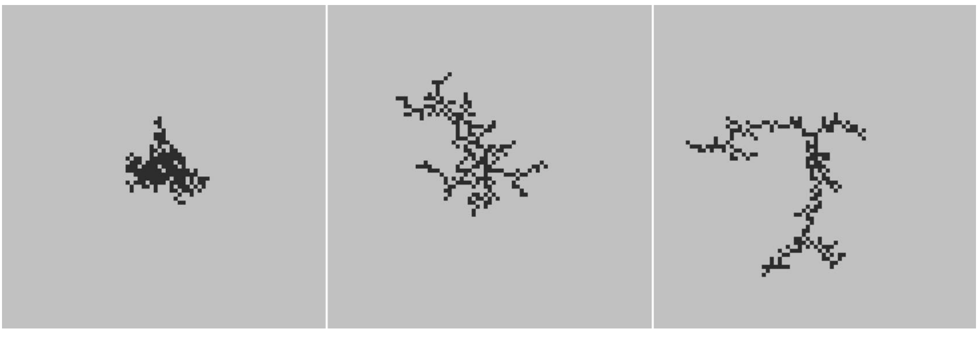
$$s_a^A(i,j) = \begin{cases} \dfrac{\exp[\beta E_a(i,j)]}{Z_a^A}(1-\epsilon), & \text{iff cell is adjacent to urban land,} \\ \dfrac{\exp[\beta E_a(i,j)]}{Z_a^A}\epsilon, & \text{else.} \end{cases} \qquad (25)$$

These small modifications are interesting because they connect empirical, theoretical, and simulation results by different research groups. For a more detailed discussion of these issues, see section 5.

4.1 Unwilling neighbors—similarities with DLA

As an example of a dynamics that generates structures similar to DLA, we define the following simple growth rule which we call the 'unwilling neighbors' rule. We find this rule to apply reasonably well to several types of land use such as (1) single-family residential and (2) low-density commercial. The rather simple rationale for this rule is as follows: (1) a common desire for single families is to live in a less-developed area and yet have access to the infrastructure of the city; (2) commercial land uses such as supermarkets with customers mainly using cars for transportation have reason to seek out land that is connected to infrastructure but less developed because of lower land prices.

This translates roughly into a rule where undeveloped land is attracting and urbanized land is repelling. However, new houses must appear adjacent to existing buildings in order to be connected to infrastructure such as streets, power, sewage, etc. In figure 6 we see the results of a series of simulations using this rule. Note that no removal occurs, only addition of new urbanized cells. The configurations visually resemble DLA configurations for large parts of the parameter space (γ and β) provided that growth can take place unhindered. The measured radial fractal dimension is around 1.5 for these regions of parameter space.



(a) (b) (c)

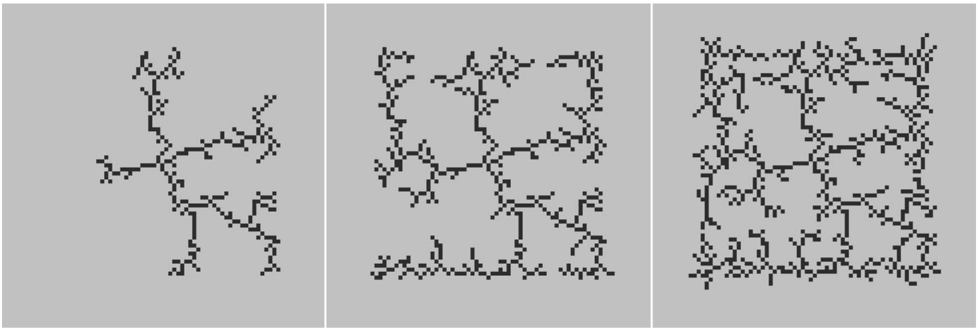
Figure 6. Dendritic, DLA-like structures. The morphology is affected by the $\beta = T^{-1}$ parameter: (a) $\beta = 0.03125$, (b) $\beta = 0.5$, (c) $\beta = 1.0$. The figure demonstrates two distinct regimes of behavior and the transition between them: structured and nonstructured.

4.2 Similarities with space-filling fractals

The logic of the 'unwilling neighbors' rule suggests that, although similar to DLA, the dynamics are rather those of a space-filling fractal. Land is developed by using a greedy algorithm that maximizes the probability to select land that is least surrounded by urban land. However, it is not until the structure begins to run out of space that the difference becomes obvious—as long as crowding does not occur, the growth rule never has to reconsider previously suboptimal areas.

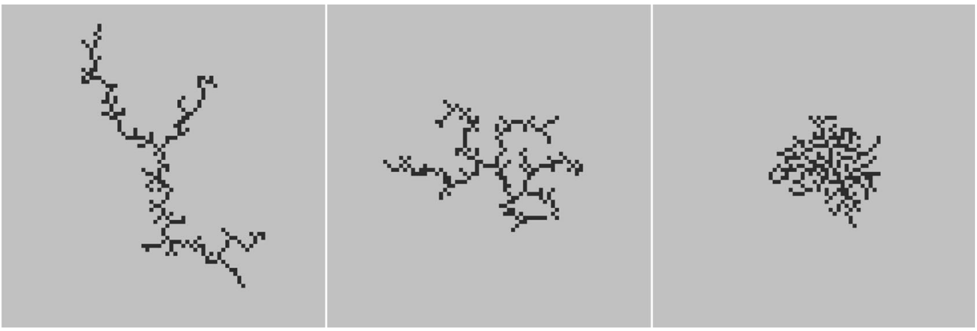
4.3 Global versus local interactions

Mean-field models have been proven capable of producing DLA structures (Muthumukar, 1983). The renormalization approach with long-range interactions used in our simulation is crucial for bringing about this DLA-like behavior (figure 7) and it cannot be reproduced using a short-range neighborhood cellular automaton. In figure 8 we see how the dentritic properties of the configurations disappear as fewer scales are employed (less global, mean-field information). The potential (energy) that drives the dynamics is depicted in color plate 3 which demonstrates why MRFs with only local interactions cannot produce DLA-like dynamics.



(a) (b) (c)

Figure 7. In this figure, the boundary conditions are such that the surrounding land has no land use (as opposed to undeveloped land use). Hence, the structure avoids the edges and thereby inherits the square shape of the lattice. In (a) the structure is not yet constrained by the grid's edges and still very much resembles a DLA, $t = 250$. In (b) the structure is beginning to be constrained and starts to grow parallel to the edges of the map, $t = 500$. In (c) the structure also begins to grow back on itself, all the time locally minimizing penalty, $t = 750$.



(a) (b) (c)

Figure 8. Removal of the influence from successively lower scales until only the nearest-neighbor interactions are left: (a) all scales used, (b) top scale ignored, (c) two top scales ignored.

4.4 Realistic urban settlements

Here we present results from simulations using multiple land-use classes with arguably realistic interactions. This cell configuration will be referred to as 'Rockville' (color plate 4 and appendix B). Suitability data calculated from a generated topographical map are used. Because urban growth is a dynamical system, the impact of suitability data is quite profound: even minute differences in development, initially brought about by a small difference in suitability, can be reinforced over time and become a major feature in the future configuration.

The simulation that is shown here uses an 81×81 lattice. With the $(1-\epsilon)$ rule as defined in section 4, the result is that the majority of all growth takes place along the edge between urban and rural land. However, new centers of growth occur constantly and serve as precursors for the spread of developed land. The dynamic land uses are: residential, commercial/industry/wholesale, and infrastructure/minor buildings represented by the colors red, grey, and light grey, respectively. Undeveloped land is the ray-traced green surface and the only static land-use class used is limited-access roads, represented by black.

5 Other urban simulations

The multiactivity simulation introduced here can be transformed, through a few minor changes, into, or more correctly approximated to, the urban evolution simulation originally developed by White et al (1997). Most importantly, a high β value is needed as it corresponds to a decrease in noise through a lowering of the temperature. This ensures that the most desirable activities get their fitness values significantly increased. To maintain a reasonable noise level in the dynamics, noise now has to be added to these selection values before a global selection occurs. By introducing a cutoff to the recursively defined levels of interaction the interaction radius can be limited to a small area around each cell as in the White et al simulation. To make actual predictions using the current simulation we also need to introduce suitabilities as defined in equation (12) derived from zoning and topography which is already included in the simulation developed by White et al. An example of the use of suitabilities to illustrate realistic urban dynamics in our current simulation is demonstrated in section (4.4) and color plate 4.

The urban growth simulation originally developed by Clarke et al (1997) is also based on a microscopic selection dynamics where existing settlements and transportation infrastructure enhance the desirability for urbanization. It has a single urbanized land-use type which our model can also be defined to have. The dynamics of this simulation are defined through a four-step process consisting of: (1) spontaneous growth, (2) new spreading centers, (3) edge growth, and (4) road-influenced growth (Candau et al, 2000). The two main differences between our current approach and the Clarke et al approach are that we use a longer range (explicit) potential (energy) function. Second, our approach in its simplest form does not account for a preferred edge growth, which is very important to obtain realistic urban morphology. We have to introduce this edge effect as an additional condition. Obviously, the Clarke et al simulation has an implicit 'energy' function which promotes growth at the edges of existing urbanized cell clusters and only a uniformly shallow energy potential away from these urbanized clusters. The concept of layers in the Clarke model is almost analogous to the concept of static land-use classes in our model.

Noting that urbanization often takes place on the edges of an urban area, cluster models from statistical physics have been applied to explain urban morphology. In particular, the DLA model (Batty and Longley, 1994) and models of correlated percolation (Makse et al, 1995; 1998) have been successful in replicating some of the general behavior of urban evolution. The relation between r, the radius from the 'central business district', and population density ρ has been found (Clark, 1951) to be

$$\rho(r) = \rho_0 \exp(-\lambda r) . \tag{26}$$

The DLA model generates a branching structure with a fractal geometry that predicts that the relationship should be a power law,

$$\rho(r) \sim r^{D-2} . \tag{27}$$

Thus, it does not produce the correct geometry in itself. The correlated percolation model, however, corrects this by using a density gradient that conforms to equation (26).

Another example of such an edge process is developed by Schweizer and Schimansky-Geier (1994) in their 'active walker' approach. Both the DLA dynamics and the active walker processes are fundamentally a diffusion aggregation process (either passive or active) where small particles of 'free' built structures diffuse around and between existing built areas and eventually attach themselves to existing built areas once the two 'touch' each other. Despite the apparent lack of correspondence between the real microdynamics of urban growth based on desirabilities, pricing, roads, etc, this diffusion aggregation process generates urban aggregates with several macroproperties also observed in real urban areas such as the morphology and the fractal dimension.

The model developed in this paper also bears some resemblance to the master-equation-based sociodynamics approach (Weidlich, 1995; 2000) in that both generate spatial structure in response to a field generated by the structure itself, and both are forced by exogenously determined growth. But the sociodynamic model contains a global structuring term, so that one of the determinants of land use is distance from the previously specified center of the urban structure; in this sense, the structure is to a certain degree imposed rather than generated, although another model is able to generate organizing centers.

The concept of using a hierarchy of scales for describing urbanization is well established in geography through central place theory (Christaller, 1933; White, 1977; 1978), and renormalization is a standard method in statistical physics. Spatial interaction models have been used extensively in the past for social modeling and introduced the concept of using interactions between activities on a lattice in models of urban growth and transportation (Batty, 1991; Golledge and Stimson, 1997; Helbing, 1995; Nagel et al, 1997; Transims, 1998; Weidlich, 1995; 2000; Wilson, 1970; Yamins et al, 2002). These are in a sense similar to the approach developed here—for example, both assume that spatial structure and spatial interaction are mutually determined. However, the emphasis was primarily on predicting interzonal flows rather than on achieving an understanding of spatial structure.

Finally we should mention a simulation of large-scale systems of settlements that is proposed by Zanette and Manrubia (1997; 1998). They use a reaction–diffusion process on a 2D lattice. Results are consistent with the well-known scaling law that states that frequency as a function of city size follows a power law,

$$\mathrm{f}(n) \propto n^{-r} , \tag{28}$$

where n is the number of inhabitants and $r \approx 2$. The reasoning behind this intermittency model stems from the theory of self-organized criticality (Bak, 1996). Dynamical systems with an external driving force, such as Bak's classical sand-pile model, often exhibit a power-law scaling of event sizes (such as avalanches in the case of the sand pile).

6 Conclusion

Our urban growth simulation is based on the multiple-land-use concepts originally developed by White et al (1997). However, it takes into account land-use interactions over the entire lattice and uses a modified Markov random field as the (energy) potential function. Instead of letting the introduction of new land use be governed by purely local, independent probabilities as in an ordinary MRF, we use a global selection similar to selection on a fitness landscape.

In a single land-use version of the simulation, we initially discuss clustering versus nonclustering dynamics as a function of what corresponds to temperature and pressure and demonstrate how these qualitatively different dynamics are analogous to different physical states of aggregation (liquid and gas). In a multiple land-use model we change the land-use – land-use interaction functions and demonstrate how multiple phase separations and phase structuring occur between the different land-use classes. These findings are discussed in the context of observed urban dynamics, and examples are used to illustrate their validity.

We then investigate how slight modifications to the MRF formulation brings us close to previously defined urban growth models. The addition of a rule that requires urbanized cells to appear adjacent to already urbanized cells generates behavior very similar to that of a diffusion-limited aggregation model which has mainly been discussed as a possible model of urban growth by the physics community (Batty, 1991; Batty and Longley, 1994; Makse et al, 1995; 1998; Stanley et al, 1996). As a more general formulation of this modification we allow a small urbanization ϵ away from the already urbanized areas and still have most of the urbanization at the edges of already urbanized areas $(1-\epsilon)$. Exclusive edge growth thus corresponds to $\epsilon = 0$, using this formulation. On the other hand, using $\epsilon > 0$ we formulate a model that is very similar to the urban growth models developed by Clarke et al (1997) and capable of generating growth patterns reminiscent of urban sprawl. With this approach we demonstrate the ability of our simulation to produce state configurations similar to observed cities.

The importance of aggregational processes in models of urban growth has been noted by several researchers (Batty and Longley, 1994; Makse et al, 1998; Witten and Sander, 1983). Although this is one important aspect, another equally important is the interaction mode of CA where land parcels are dynamically updated depending on their neighborhood. The analogies that can be made between these two approaches and the workings of the real-world system are very clear and intuitive. Nevertheless, a classical CA is not even in principle capable of reproducing the spatial dynamics of a DLA and a DLA lacks many obvious aspects of realism that are captured in the CA approach. Our model bridges this gap without a dramatic increase in computational complexity and thus lets us combine the benefits of both approaches (Muthumukar, 1983).

Acknowledgements. Because of its cross-scientific nature, formulating this paper held a lot of special challenges. We are indebted to numerous people for making the paper possible by reviewing, discussing, and providing feedback and additions to its content as well as its formulation. Especially, we wish to thank the following: Michael Brown, Jeanette Candau, Keith Clarke, Helen Couclelis, Noah Goldstein, Grant Heiken, Xaiohang Liu, Martin Nilsson, Mats Nordahl, Tim Robinson, and Larry Winther. This work was supported by The Urban Security Initiative and STB Student program, both LANL as well as the Santa Fe Institute.

References

Alonso W, 1964 *Location and Land Use* (Harvard University Press, Cambridge, MA)
Bak P, 1996 *How Nature Works* (Springer, New York)
Batty M, 1976 *Urban Modelling: Algorithms, Calibrations, Predictions* (Cambridge University Press, Cambridge)

Batty M, 1991, "Generating urban forms from diffusive growth" *Environment and Planning A* **23** 511 – 544
Batty M, Longley P, 1994 *Fractal Cities: A Geometry of Form and Function* (Academic Press, London)
Batty M, Xie Y, 1994, "From cells to cities" *Environment and Planning B: Planning and Design* **21** S31 – S48
Benenson I, Portugali J, 1997, "Agent-based simulations of a city dynamics in a GIS environment" *Spatial Information Theory: A Theoretical Basis for GIS* Eds S C Hirtle, A U Frank (Springer, Berlin) pp 501 – 502
Benguigui L, 1995, "A new aggregation model: application to town growth" *Physica A* **219** 13 – 26
Benguigui L, Daoud M, 1991, "Is the suburban railway system a fractal?" *Geographical Analysis* **23** 362 – 368
Broadbent T A, 1971, "A hierarchical interaction-allocation model for a two-level spatial system" *Regional Studies* **5** 23 – 27
Candau J, Rasmussen S, Clarke K C, 2000, "A coupled cellular automaton model for land use cover dynamics", in *Proceedings of 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4) Banff, Alberta, Canada, September 2 – 8 2000*
Christaller W, 1933 *Die Zentralen Orte en Süddeutschland* [Central places in Southern Germany] (Gustav Fischer, Jena)
Clark C, 1951, "Urban population densities" *Journal of the Royal Statistical Society A* **114** 490 – 496
Clarke K C, Gaydos L, 1998, "Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore" *International Journal of Geographical Information Science* **12** 699 – 714
Clarke K C, Hoppen S, Gaydos L, 1997, "A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area" *Environment and Planning B: Planning and Design* **24** 247 – 261
Couclelis H, 1997, "From cellular automata to urban models: new principles for model development and implementation" *Environment and Planning B: Planning and Design* **24** 165 – 174
Engelen G, White R, Uljee I, 1997, "Integrating constrained cellular automata models, GIS and decision support tools for urban planning and policy-making", in *Decision Support Systems in Urban Planning* Ed. H Timmermans (E & F N Spon, London) pp 125 – 155
Golledge R G, Stimson R J, 1997 *Spatial Behavior: A Geographic Perspective* (Guilford Press, New York)
Helbing D, 1995 *Quantitative Sociodynamics: Stochastic Methods and Models of Social Interaction Processes* (Kluwer Academic, Dordrecht)
Kauffman S, 1993 *The Origins of Order* (Oxford University Press, Oxford)
Kindermann R, Snell J L, 1980, "Markov random fields and their applications" *Contemporary Mathematics, American Mathematical Society* **1** .....
Langton C, 1992, "Life at the edge of chaos", in *Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems. Santa Fe Institute Studies in the Science of Complexity, Volume 10* (Addison-Wesley, Reading, MA) pp 41 – 92
Makse H A, Havlin S, Stanley H E, 1995, "Modelling urban growth patterns" *Nature* **377** 608 – 612
Makse H A, de Andrade J S, Batty M, Havlin S, Stanley H E, 1998, "Modeling urban growth patterns with correlated percolation" *Physics Review E* **477** 608 – 612
Muthumukar M, 1983, "Mean field theory for diffusion-limited cluster formation" *Physics Review Letters* **50** 839 – 842
Nagel K, Rasmussen S, Barrett C L, 1997, "Network traffic as a self-organized critical phenomenon", in *Self-organization of Complex Structures: From Individual to Collective Dynamics* Ed. F Schweitzer (Gordon and Breach, London) pp 579 – 593
Papageorgiou Y, 1990 *The Isolated City State: An Economic Geography of Urban Spatial Structure* (Routledge, London)
Portugali J, 2000 *Self-organization and the City* (Springer, Berlin)
Sanders L, Pumain D, Mathian H, Guérin-Pace F, Bura S, 1997, "SIMPOP: a multiagent system for the study of urbanism" *Environment and Planning B: Planning and Design* **24** 287 – 305
Schweitzer F, Schimansky-Geier L, 1994, "Clustering of active walkers in a two-component system" *Physica A* **206** 359 – 379
Stanley H E, Amaral L A N, Buldyrev S V, Goldberger A L, Havlin S, Leschhorn H, Maass P, Makse H A, Peng C-K, Salinger M A, Stanley M H R, Viswanathan G M, 1996, "Scaling and universality in animate and inanimate systems" *Physica A* **231** 20 – 48

Takeyama M, 1996, "Geocellular: a general platform for dynamic spatial simulation", in *Artificial Worlds and Urban Studies* Eds E Besussi, A Cecchini (Dipartimento di analisi economica e sociale del territorio, Venice) pp 347 – 364

Transims, 1998, http://transims.tsasa.lanl.gov/

Weidlich W, 1995, "From fast to slow processes in the evolution of urban and regional settlement structures" *Discrete Dynamics in Nature and Society* **3** 137 – 148

Weidlich W, 2000 *Sociodynamics: A Systematic Approach to Mathematical Modelling in the Social Sciences* (Harwood Academic, London)

White R, 1977, "Dynamic central place theory: results of a simulation approach" *Geographical Analysis* **9** 227 – 243

White R, 1978, "The simulation of central place dynamics: two sector systems and the rank-size distribution" *Geographical Analysis* **10** 201 – 208

White R, Engelen G, 1993, "Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns" *Environment and Planning A* **25** 1175 – 1199

White R, Engelen G, 1997, "Multi-scale spatial modelling of self-organizing urban systems", in *Self-organization of Complex Structures: From Individual to Collective Dynamics* Ed. F Schweitzer (Gordon and Breach, London) pp 519 – 535

White R, Engelen G, Uljee I, 1997, "The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics" *Environment and Planning B: Planning and Design* **24** 323 – 343

Wilson A G, 1970 *Entropy in Urban and Regional Modeling* (Pion, London)

Witten T A, Sander L M, 1983, "Diffusion-limited aggregation" *Physics Review B* **27** 5686 – 5697

Wu F, 1998a, "An experiment on the generic polycentricity of urban growth in a cellular automatic city" *Environment and Planning B: Planning and Design* **25** 731 – 752

Wu F, 1998b, "Simland: a prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rule" *International Journal of Geographical Information Science* **12** 63 – 82

Xie Y, 1996, "A generalized model for cellular urban dynamics" *Geographical Analysis* **28** 350 – 373

Yamins D, Rasmussen S, Fogel D, 2002, "Growing urban roads" *Networks and Spatial Economics* forthcoming

Zanette D H, Manrubia S C, 1997, "Role of intermittency in urban development: a model of large-scale city formation" *Physics Review Letters* **79** 523 – 526

Zanette D H, Manrubia S C, 1998, "Intermittency model for urban development" *Physics Review E* **58** 295 – 302

Appendix A—selection algorithm

To clarify the selection algorithm, let us consider the following example. We have a 2×2 grid S of potentials (a very small map) for a single land use a with $n = 2$: $S = \{\{2, 3\}, \{7, 3\}\}$. Then, $\Sigma_a^R = 15$, and the partial sums are $\Sigma^* = \{\{2, 5\}\,\{12, 15\}\}$. The range in which we pick our random number r would be $[0, 15)$, so if we pick an arbitrary number, say 11, we find that $5 \leqslant 11 < 12$ for $k = 2$. This is the cell we transform to the undeveloped land use.

By repeatedly picking random numbers and skipping over already selected cells it is possible to reuse equation (19) many times and thus remove exactly the desired number of new activities in each update.

Appendix B—parameters used for 'Rockville'

We have used a lattice size of 81×81 and parameter values $\beta = 3.0$, $\gamma = 0.9$, and $\epsilon = 0.001$. The following set of land uses: $C = \{$residential, commerce/industry, mixed, undeveloped, limited-access highway$\}$. Of these the first three grow dynamically. The index of each land use in C will be used hereafter to identify them. Each time t the following number of each land-use type was added to the configuration at time $t - 1$ on average: 0.5, 1, 0.5.

Interactions between the land uses are defined as follows [see equation (8)]:

$$
\begin{aligned}
A_{00} &= \{2, -2, 2, 0\},\ A_{01} = \{-1, -1, 2, 1\},\ A_{02} = \{2, 2, 0, 0\},\ A_{03} = \{0, 0, 0, 0\},\\
&\quad A_{04} = \{-1, 1, 1, 1\},\\
A_{10} &= \{0, 0, 1, 1\},\ A_{11} = \{1, 1, -1, 2\},\ A_{12} = \{1, 1, 0, 0\},\ A_{13} = \{0, 0, 0, 0\},\\
&\quad A_{14} = \{1, 2, 2, 2\},\\
A_{20} &= \{-0.5, 0, 2, 2\},\ A_{21} = \{2, 2, 0, 0\},\ A_{22} = \{-4, -4, 0, 2\},\\
&\quad A_{23} = \{0, 0.5, 0.5, 1\},\ A_{24} = \{-2, 0, 0, 0\}.
\end{aligned}
$$